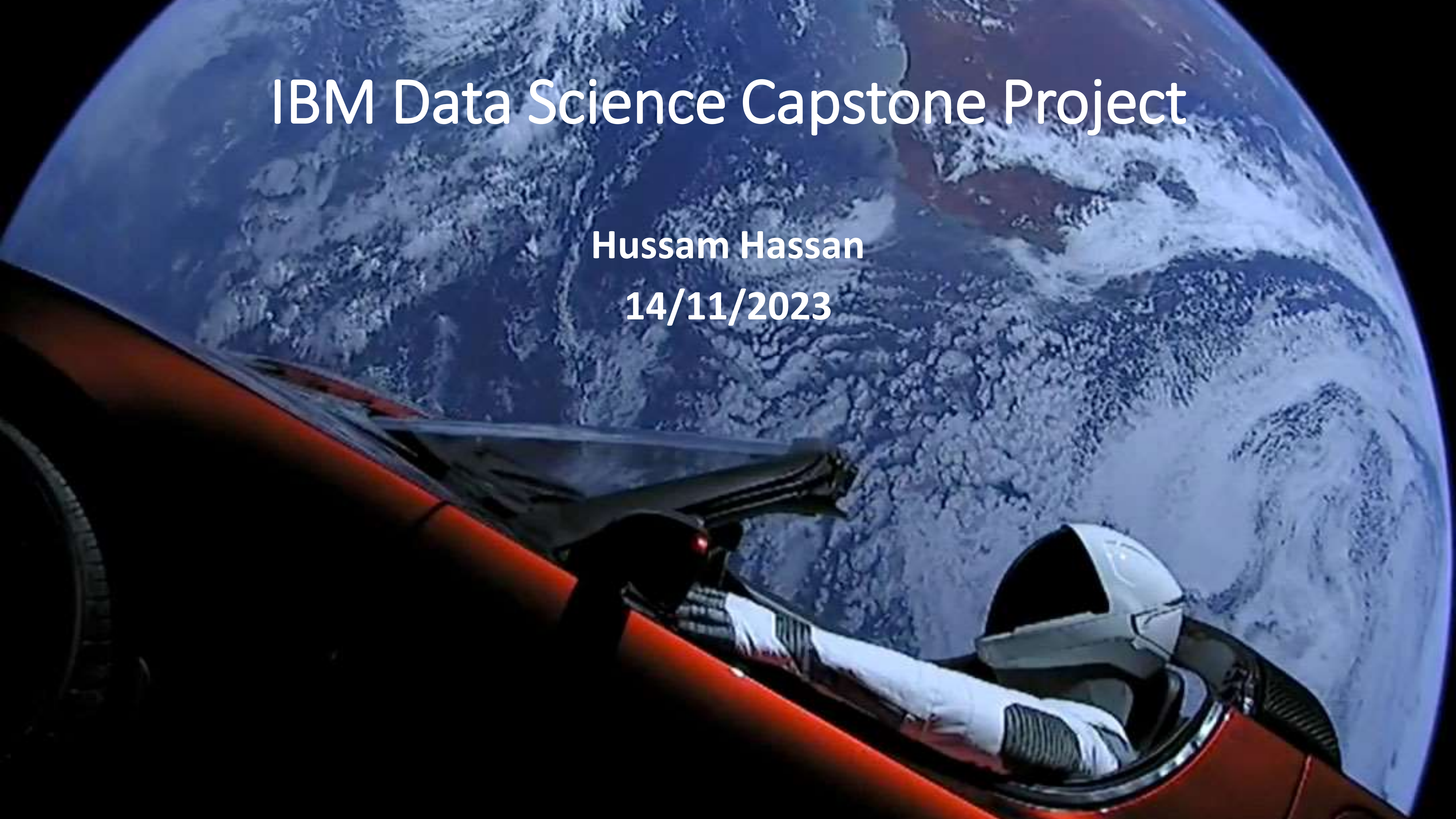


IBM Data Science Capstone Project

Hussam Hassan

14/11/2023

Outline

- Executive Summary
- Introduction
- Methodology
- Results
  - Visualization – Charts
  - Dashboard
- Discussion
  - Findings & Implications
- Conclusion

Executive Summary

- Data Collection:
  - Data Collection using an API
  - Data Collection with Web Scraping
- Data Wrangling
- EDA
  - EDA with SQL
  - EDA with Visualizations
- Interactive Analytics:
  - Interactive Analytics with Folium
  - Interactive Analytics with Plotly Dash
- Predictive Analytics

Introduction

SpaceX is a rocket manufacturing company. Their rocket falcon 9 is advertised at $65,000,000 while other manufactures advertise their rockets at $165,000,000. SpaceX rockets are much cheaper because they can reuse the first stage. The problem with Falcon 9 is that not all rockets land successfully. So we here at SpaceY would like to determine the price of Falcon 9 by predicting if the rocket would land successfully or not.

Methodology

Methodology

- A SpaceX API and Web Scrapping were used for Data Collection
- Data Wrangling was implemented using Pandas
- EDA was performed with the following libraries:
  - SQL
  - Matplotlib
- The following libraries were used for Interactive Analytics:
  - Folium
  - Plotly
  - Dash
- Predictive Analytics was performed by using Scikit-learn

Data Collection

In every data science project the data needs to be collected first. We have implemented two methods in this project, which were Data Collection using an API and Web Scrapping. Using an API of course was simple and straight forward while Web Scrapping was a bit more complicated and needed more preprocessing techniques in order to get to our final result

Data Collection

SpaceX API

- Data was requested and parsed, and then immediately was converted to a Pandas dataframe
- The data was then filtered to include the Falcon 9 launches

Web Scrapping

- A request was sent to get the Falcon9 launch wiki page
- All columns and tables were extracted from the HTML table header
- A dataframe was then created by parsing the launch HTML table

Data Wrangling

- In this section, Pandas was used to perform EDA. We first calculated the number of launches on each site, the number and occurrence of each orbit and also the number and occurrence of mission outcomes for each orbit.
- These calculations helped us create a landing outcome label, which will be useful for predictive analytics.
- Finally the results were then exported to a csv file

EDA with SQL and Visualization

EDA with SQL

- SQL was used to query data in order to gain meaningful insights of the data we have at hand.

EDA with Visualization

- Scatter plots were created to show the correlation between different variables.
- These results were then used to determine the training labels.

Interactive Analytics

Interactive Analytics with Folium

- A geospatial map of all the launch sites was created to gain more insights on the data collected.

Interactive Analytics with Plotly and Dash

- A web based application was created to plot a pie chart and a scatter plot
- The pie chart revealed the success rate for every launch site.
- The scatter plot revealed the success rate according to the payload mass carried.

Predictive Analytics

- The data was first split into training and test labels
- The algorithms used for training were obtained from the scikit-learn library
- Finally all the models were compared and the model with the best performance was chosen for predictive analytics.

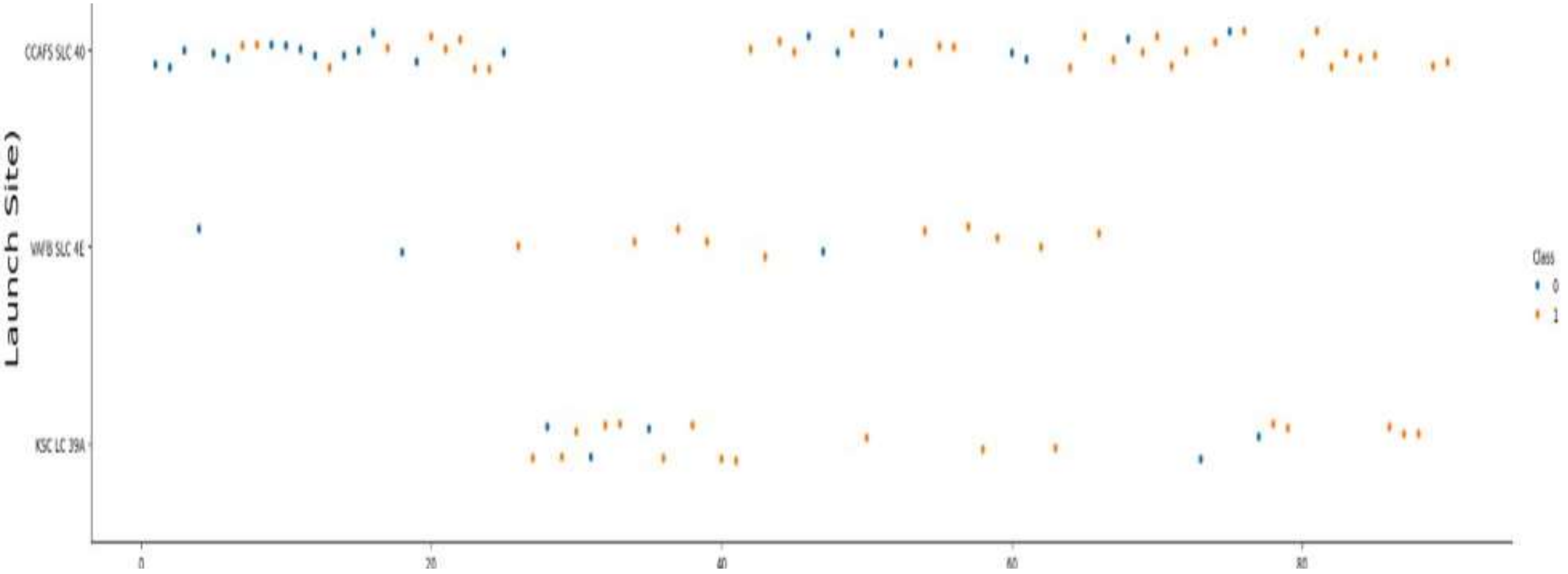
Results

EDA Results Visualization

- Scatter plots:
    - Flight Number vs Launch Site
    - Payload Mass vs Launch Site
    - Flight number vs Orbit
    - Payload Mass vs Orbit
- Bar Chart of Success Rate for each Orbit
- Line Plot of Succes rate through time

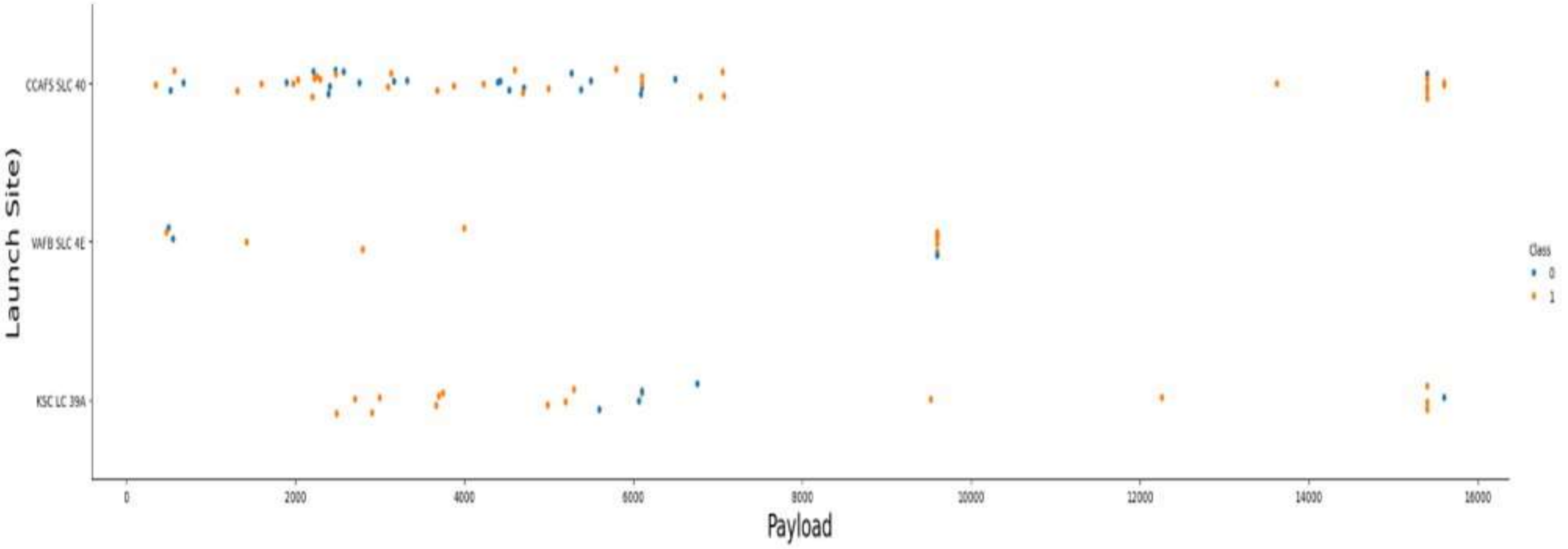
Flight Number vs Launch Site

In this chart we can see that the success rate increased as the number of flights increased.



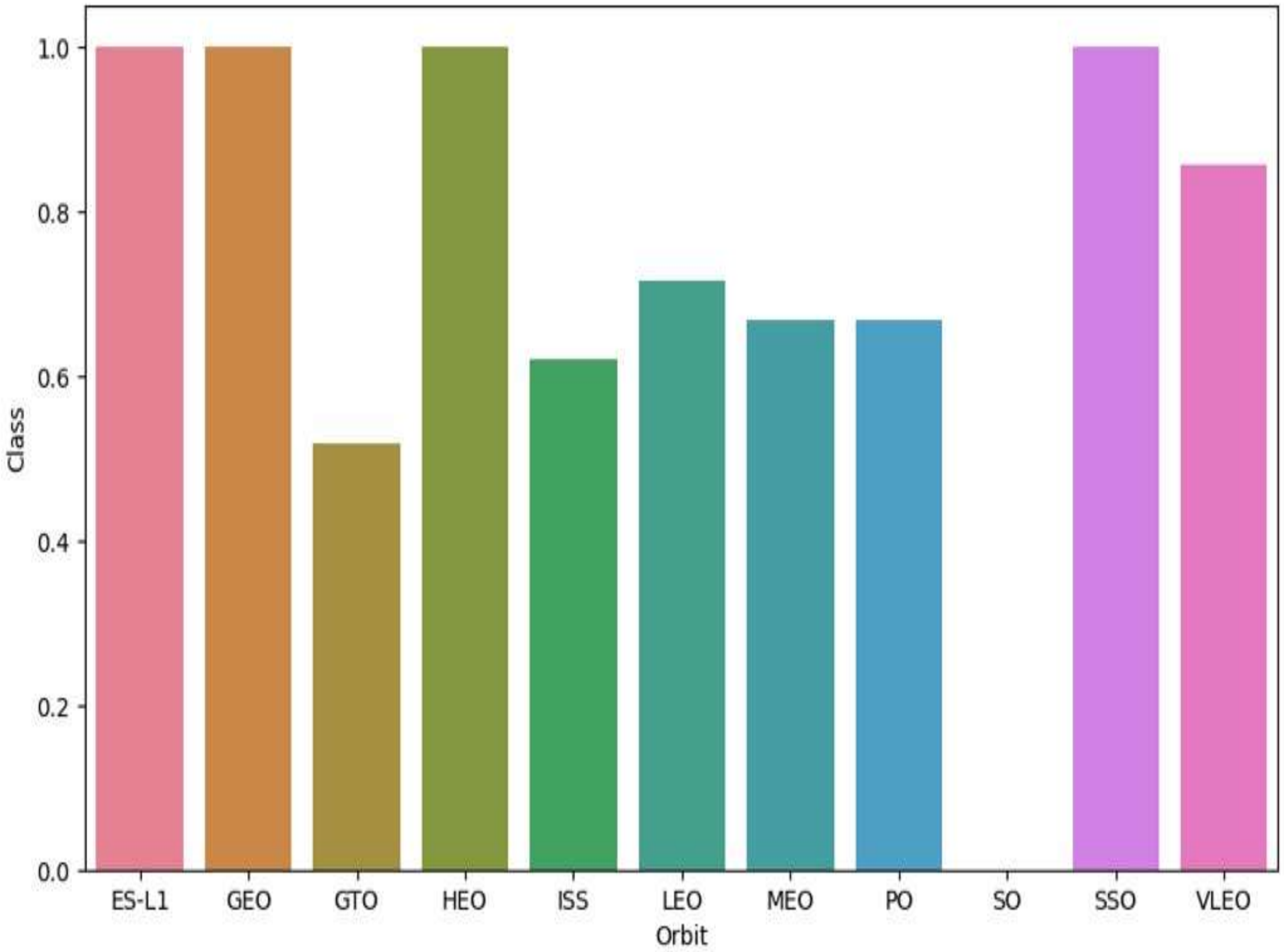
Payload Mass vs Launch Site

The number of number of success launches were fine in both small and heavy payloads but we can also see that on the VAFB SLC-4E launch site there were no rockets with heavy loads launched



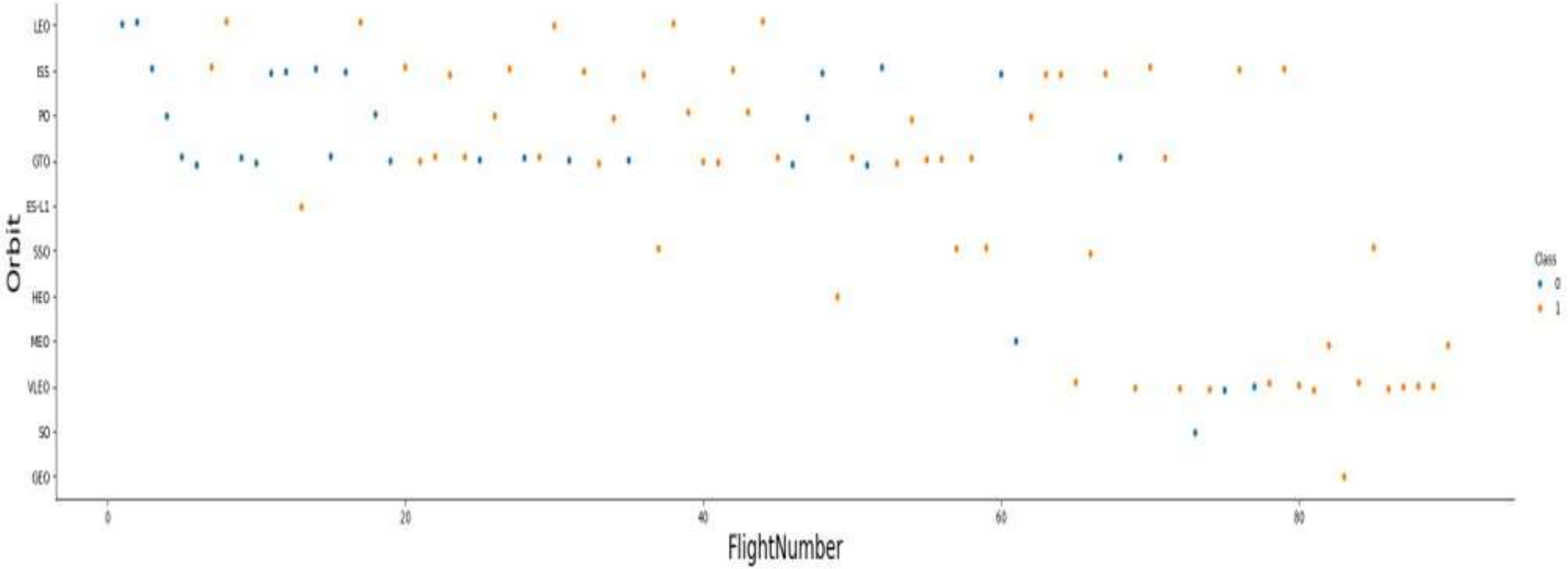
Succes rate for each Orbit

- The following Orbits had a success rate of over 80%:
  - ES-LI
  - GEO
  - HEO
  - SSO



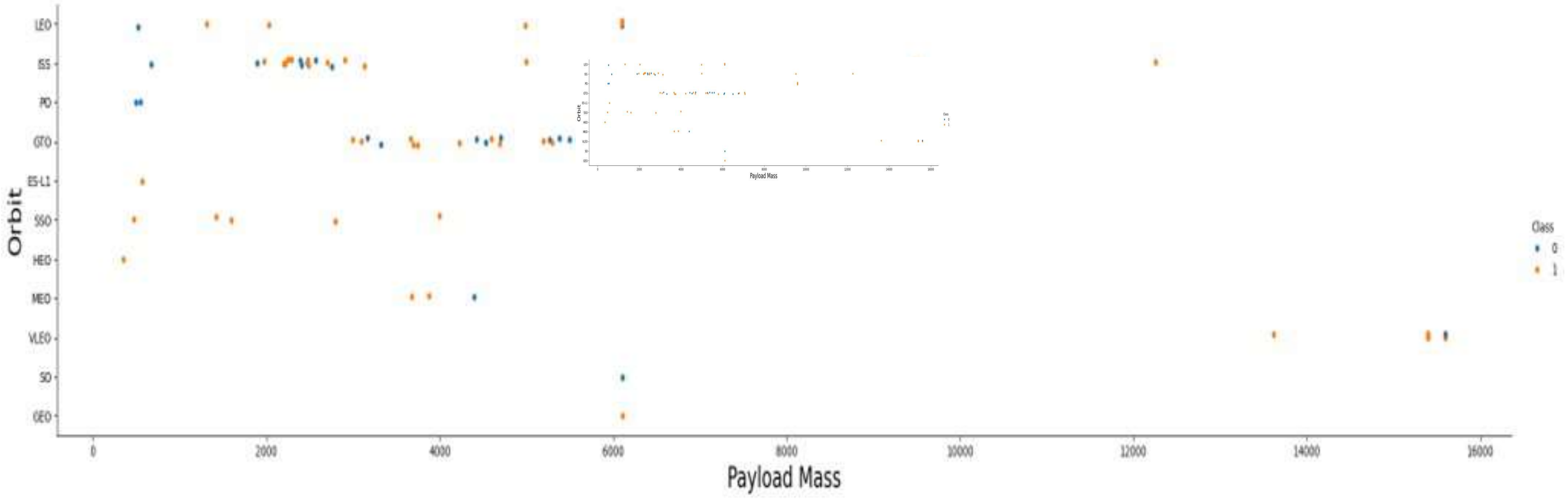
Flight Number vs Orbit

We can see that in the LEO orbit the Success appears related to the number of flights; on the other hand, there seems to be no relationship between flight number when in GTO orbit.

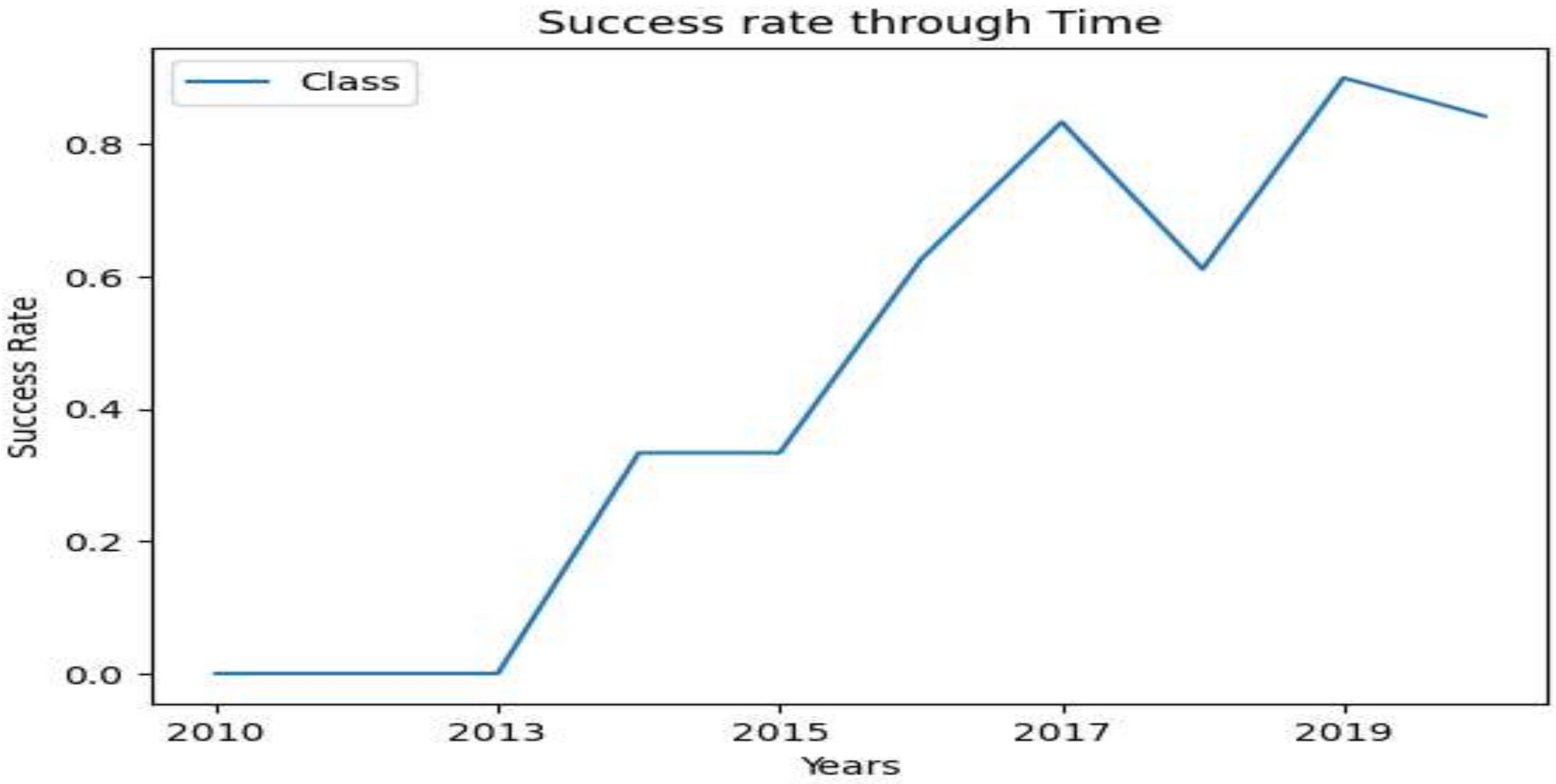


Payload Mass vs Orbit

The successful landing or positive landing rate for heavy payloads are more for Polar,LEO and ISS. However for GTO we cannot distinguish this well as both positive landing rate and negative landing(unsuccessful mission) are both there here



Success Rate (Years)



EDA RESULTS SQL

- In this section we gained access to some interesting inights by querying data through SQL

EDA RESULTS SQL

All Launch Sites

Done.

[9]:

Launch_Site
CCAFS LC-40
VAFB SLC-4E
KSC LC-39A
CCAFS SLC-40

Average payload mass carried by booster version F9 v1.1

Done.

[12]:

Average
2928.4

Total payload mass carried by boosters launched by NASA (CRS)

Done.

[11]:

Sum
45596

Names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000

[16]:

Booster_Version
F9 FT B1022
F9 FT B1026
F9 FT B1021.2
F9 FT B1031.2

Date when the first successful landing outcome in ground pad was achieved.

[14]:

MIN(Date)
2015-12-22

EDA RESULTS SQL

Names of the booster versions which have carried the maximum payload mass

[26]:

Booster_Version
F9 B5 B1048.4
F9 B5 B1049.4
F9 B5 B1051.3
F9 B5 B1056.4
F9 B5 B1048.5
F9 B5 B1051.4
F9 B5 B1049.5
F9 B5 B1060.2
F9 B5 B1058.3
F9 B5 B1051.6
F9 B5 B1060.3
F9 B5 B1049.7

Month names, failure landing outcomes in drone ship ,booster versions, launch site for the months in year 2015.

[27]:

Month	Landing_Outcome	Booster_Version	Launch_Site
20	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
/2	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40
/2	Failure (drone ship)	F9 v1.1 B1017	VAFB SLC-4E
01	Failure (drone ship)	F9 FT B1020	CCAFS LC-40
/2	Failure (drone ship)	F9 FT B1024	CCAFS LC-40

Count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order.

[28]:

Landing_Outcome	COUNT(Landing_Outcome)
Success	15
No attempt	6
Success (ground pad)	5
Success (drone ship)	5
Failure (drone ship)	4
Failure	3
Controlled (ocean)	3
Failure (parachute)	1

Interactive Analytics

Interactive Analytics with Folium

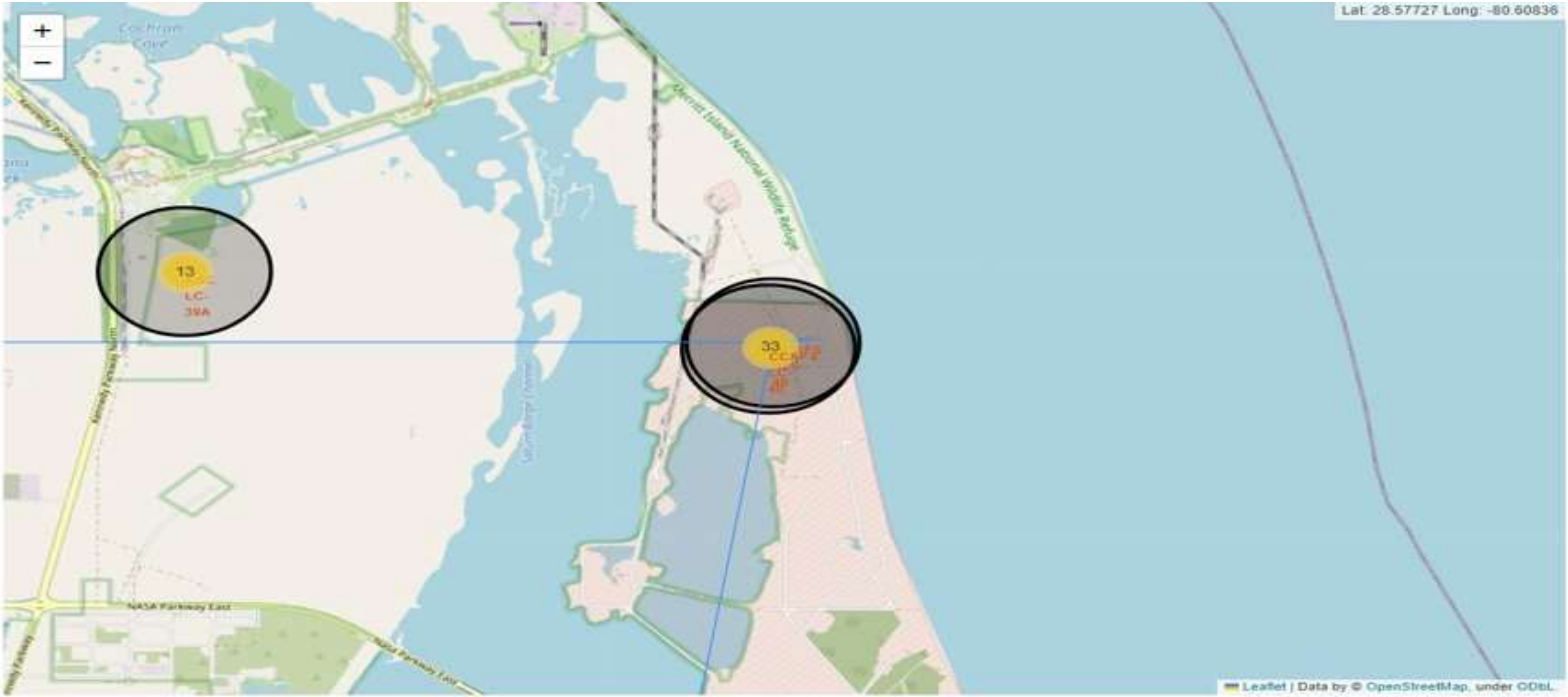
- The results obtained from the geo-spatial data did not help us a lot with our goal but provided us with some interesting information. All launch site sites were close to a railway or a highway for logistic purposes. They were also in close proximity to cities to provide accommodation for employees.

Interactive Analytics with Folium

[16]:

Interactive Analytics with Folium

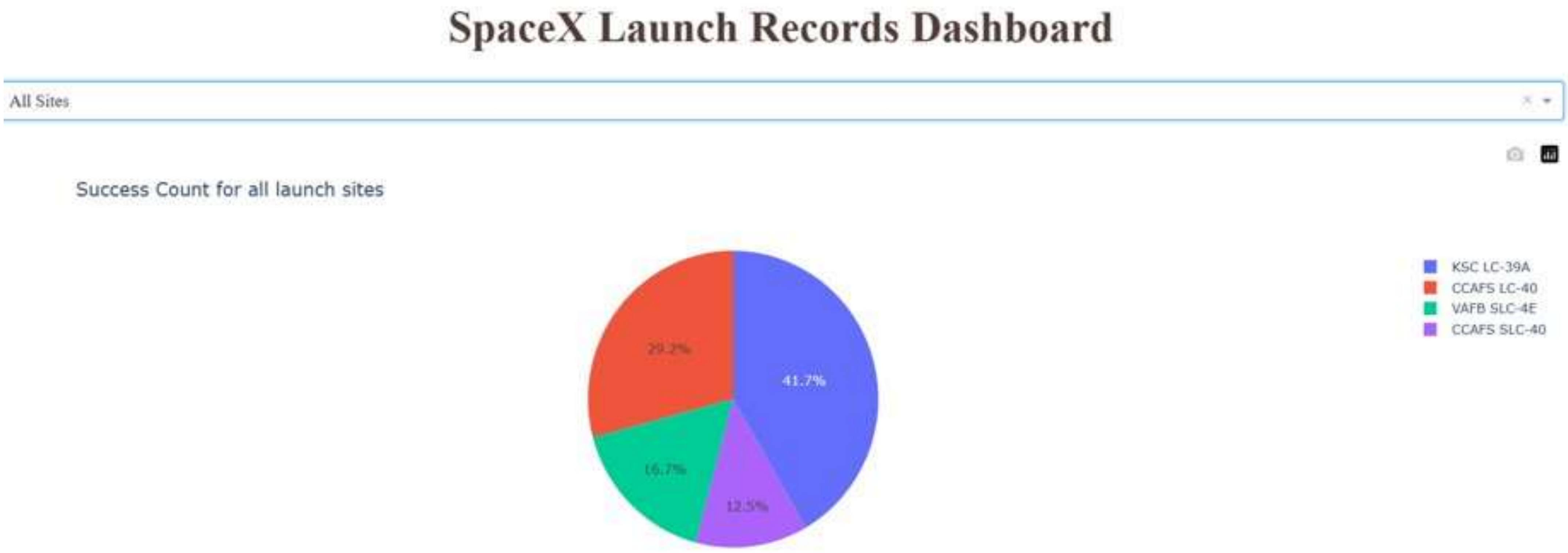
[25]:



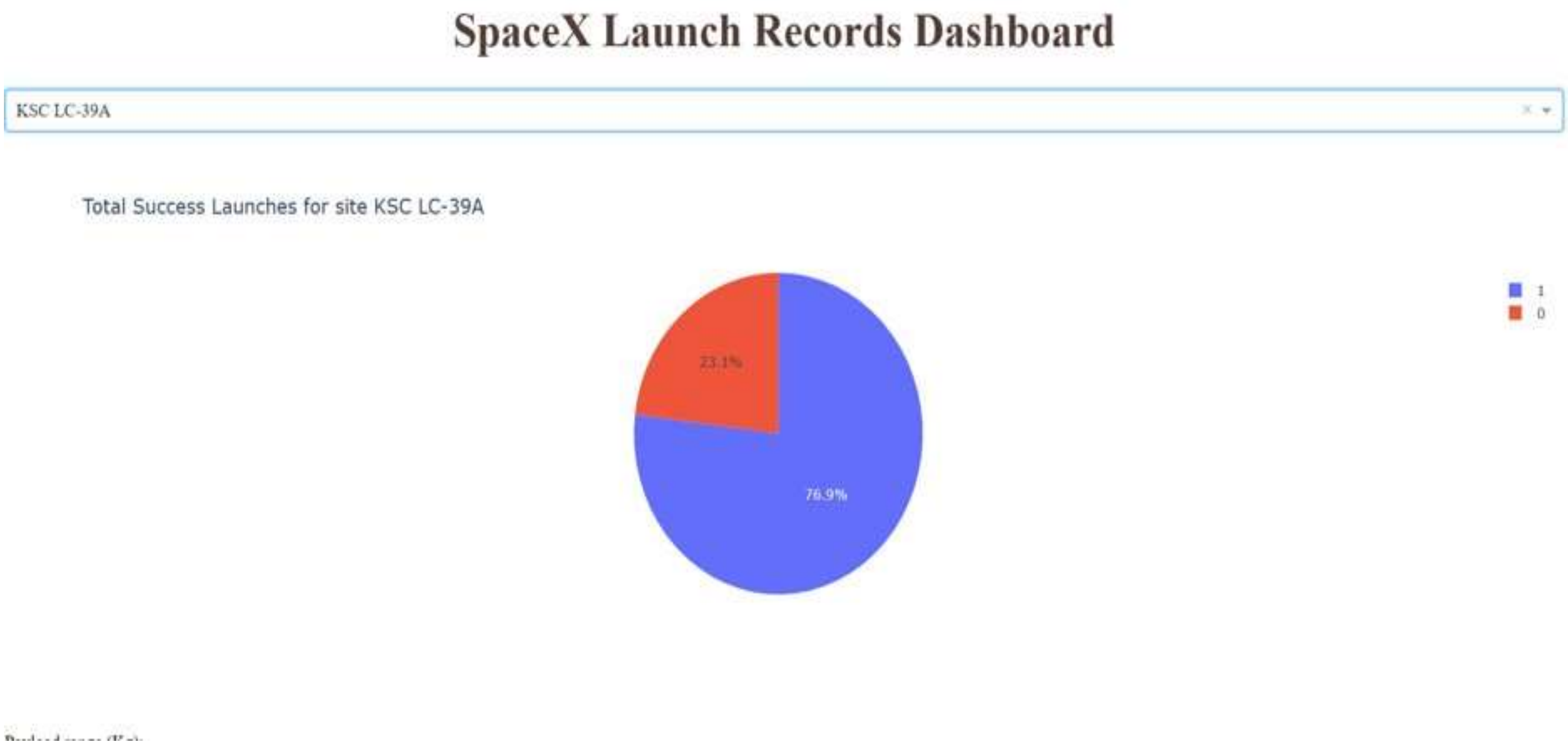
Interactive Analytics with Plotly and Dash

- The Dashboards helped us obtain the following results:
  - The KSC LC-39A had both the highest success rate and highest number of successful launches.
  - The payload mass range between 2000 and 4000 had the highest launch success rate
  - The payload mass range between 4000 and 6000 had the lowest success rate
  - The FT Booster was the version with the highest success rate

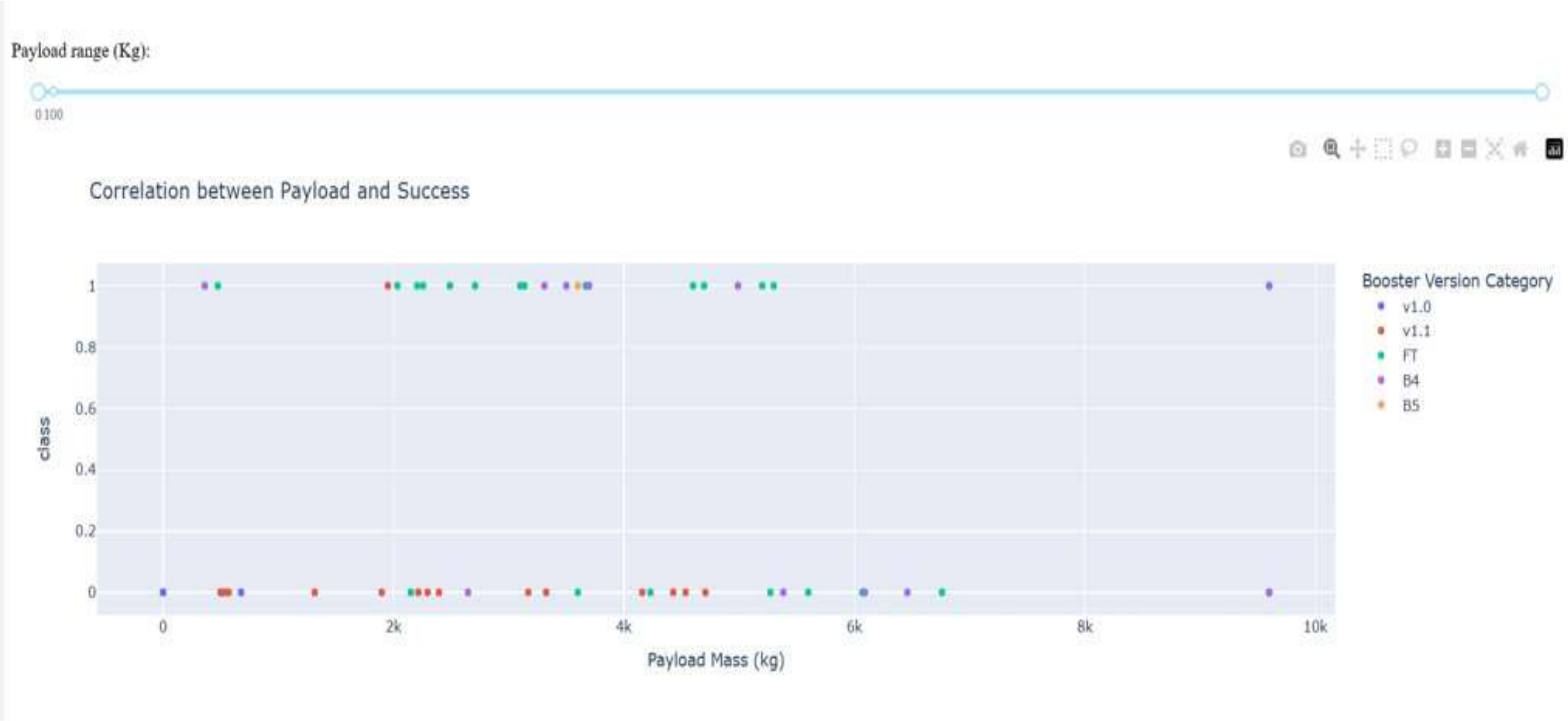
Interactive Analytics with Plotly and Dash



Interactive Analytics with Plotly and Dash



Interactive Analytics with Plotly and Dash



Predictive Analytics

Predictive Analytics

- Four different machine learning algorithms were implemented and the compared to see which would yield the best result:
  - Logistic Regression
  - Support Vector Machines
  - Decision Trees
  - K Nearest Neighbours
- The Decision Trees had the highest score while the others yielded the same results.

Conclusion

- According to the results SpaceX has done a really good job in improving their successful launches.
- Better results could have been achieved by using more data and feeding it into a more advanced machine learning algorithm such as a Neural Network.